

MULTILEVEL APPROACH FOR SIGNAL RESTORATION PROBLEMS WITH TOEPLITZ MATRICES*

MALENA I. ESPAÑOL^{†‡} AND MISHA E. KILMER[†]

Abstract. We present a multilevel method for discrete ill-posed problems arising from the discretization of Fredholm integral equations of the first kind. In this method, we use the Haar wavelet transform to define restriction and prolongation operators within a multigrid-type iteration. The choice of the Haar wavelet operator has the advantage of preserving matrix structure, such as Toeplitz, between grids, which can be exploited to obtain faster solvers on each level where an edge-preserving Tikhonov regularization is applied. Finally, we present results that indicate the promise of this approach for restoration of signals and images with edges.

Key words. ill-posed problems, multilevel, regularization, Haar wavelets, signal restoration

AMS subject classifications. 65F10, 65F22, 65F50

DOI. 10.1137/080715780

1. Introduction. Linear ill-posed problems in the form of Fredholm integral equations of the first kind occur frequently in the physical sciences [8]. They model inverse problems, that is, situations where hidden information is computed from external observations. A first-kind Fredholm integral equation for a one dimensional (1D) problem is as follows:

$$(1.1) \quad \int_{\alpha_1}^{\alpha_2} K(s, t)f(t)dt = g(s), \quad \beta_1 \leq s \leq \beta_2,$$

where the function K , called the *kernel*, is a known function of two variables s and t , and the right-hand side g is also known, while f is the unknown function that we wish to compute.

In particular, when the kernel $K(s, t)$ is *spatially invariant*, that is, its effect depends only on the distance between s and t , then the preceding equation represents a convolution integral, where $K(s, t) = K(s - t)$. In this case, we can say that $g(s)$ is the result of convolving $K(s)$ and $f(s)$.

The discretization of such an integral equation gives a *discrete ill-posed problem* [11], taking the form

$$(1.2) \quad Ax^{true} = b^{true},$$

where the matrix A is a discretization of the integral operator, and the vectors x^{true} and b^{true} are the discretized versions of the continuous functions f and g . Then, given A and b^{true} , the corresponding discrete inverse problem is to recover x^{true} .

The continuous problem (1.1) is generally ill-posed in the sense that small changes in g can cause arbitrarily large changes in f . Consequently, the matrix A is ill-conditioned, and its singular values decay to zero without significant spectral gap. In fact, instead of having b^{true} , we have measured data, b , that contains noise, e , due to

*Received by the editors February 15, 2008; accepted for publication (in revised form) May 20, 2009; published electronically February 5, 2010.

<http://www.siam.org/journals/sisc/32-1/71578.html>

[†]Department of Mathematics, Tufts University, Medford, MA 02155 (misha.kilmer@tufts.edu).

[‡]Current address: Graduate Aeronautical Laboratories, California Institute of Technology, Pasadena, CA 91125 (maspanol@caltech.edu).

measurement and/or approximation error. In this paper, we assume that e is white noise (i.e., the covariance matrix for e is a scaled identity matrix). Then, the exact solution of the system of equations

$$(1.3) \quad Ax = b = b^{true} + e$$

has little relationship to the noise-free solution x^{true} and is worthless. Thus, we seek an approximation to the solution, x^{true} , by solving a nearby more well-posed problem. This method is called *regularization*. Regularization methods [11] include Tikhonov regularization or early termination of certain iterative methods, such as conjugate gradient least squares (CGLS).

In this paper, we restrict ourselves to the case when $A \in \mathbb{R}^{m \times m}$, with $m = 2^k$ for some integer k . We focus on the 1D signal restoration problem, although the 2D (image restoration) problem is also discussed. In the 1D case, A acts as a *blurring* operator on the “signal” x^{true} , and b denotes the blurred noisy signal. We are particularly interested in the case when the signal has “edges” that we want to recover. Also, we will assume that the blurring function K is spatially invariant. Consequently, the matrix A is Toeplitz. A Toeplitz matrix A is one whose elements are constant along diagonals; that is, the (i, j) entry in A , t_{ij} , is given by t_{i-j} .

Multigrid (MG) methods [5] are well known as extremely efficient solvers for certain large-scale systems of equations, namely the ones resulting from the discretizations of certain classes of partial differential equations and integral equations of the second kind. These have been extensively studied in recent years. However, for ill-posed problems the classical multigrid approach is not immediately applicable. The first consideration is that vectors in the span of the singular vectors corresponding to the largest singular values are smooth, while the singular vectors corresponding to the small singular values represent “high frequency” information. Since the integral operator is a smoothing operator but the noise vector is assumed to be white, the expansion coefficients of b^{true} decay on average faster than the singular values, while the expansion coefficients of the noise vector are roughly constant in absolute value. Thus, the noise perturbs the high frequency components of the right-hand side and makes it impossible to recover x^{true} through the use of typical multigrid “smoothers” such as Jacobi or Gauss–Seidel, because they can be shown to incorporate noise into the solution immediately.

Multigrid methods have already been considered for solving ill-posed problems [6, 9, 13, 17, 19, 21, 22]. The two-level and multilevel methods presented in [9, 22] are applied directly to the regularized Tikhonov form of the problem. That is, the regularization of the problem is obtained by the starting model. The approaches presented in [6, 13, 17, 19, 21], on the other hand, present multilevel methods applied to the unregularized problem. Then, the regularization is obtained by the method itself. We present a new multilevel approach that follows the overall ideas found in [6], where the authors used a standard multigrid V-cycle scheme and standard (with respect to the PDE literature) prolongation and coarse-grid correction operators. The smoothers advocated were those such as CGLS, which are known to capture the smooth component of the signal at each level without mixing in noise (high frequency information) as long as they are terminated after a small number of iterations. However, their approach did not include a residual correction (post-smoothing) step on the upward part of the V-cycle, and rather used only interpolation to move back up to the fine grid. They advocated for very few iterations at each level (i.e., overregularization) and multiple V-cycles. As a result of this choice of smoother, in many cases, the edges

in the restored images are not well recovered. However, the work in [6] illustrates the potential for a multilevel approach to regularized signal and image deblurring.

Therefore, in the present work, we turn to the difficult task of recovering edge information in 1D signals using a multilevel approach. It is well known in the signal processing literature that one can restore edge information by using, for example, a Tikhonov-based approach where the regularization functional is a total-variation (TV) operator [23]. Another edge-preserving approach is the PPTSVD approach [12]. These approaches, although they can work very well, can be quite computationally expensive. Therefore, we develop a multilevel approach that has the capability to recover some edge information at a lesser expense than other traditional methods. The authors in [19] had the same goal in mind. Their approach differs from ours in two respects: (1) they work with a full multigrid approach rather than a traditional V-cycle with pre-smoothing, and (2) the edge preservation is obtained through use of a nonlinear edge-preserving prolongation operator, whereas our edge preservation is built in through the choice of appropriate coarse-grid and post-smoothing operators.

We present a method that has a standard multigrid V-cycle scheme but uses a different restriction operator than that in [6]. It also uses a residual correction step which actually lives on a coarse grid but regularizes over the next finer grid. In particular, we choose to move between grids using a Haar wavelet operator. We show that the choice of the Haar wavelet operator has the advantage of preserving matrix structure between grids, which can be exploited to obtain faster solvers on each level, and allows for a straightforward analysis of errors and residuals as we move through the V-cycle. For other multigrid methods for Toeplitz matrices, see [1, 2].

The structure of the paper is as follows. In section 2, we first give the outline of a generic multilevel algorithm for discrete ill-posed problems, and then describe how we will use the Haar wavelet operator to move between grids. In section 3, using some matrix and residual analysis, we design our pre-smoothing, coarsest-grid correction and so-called residual-correction strategies (post-smoothing). We address some of the computational issues in section 4. Section 5 presents a straightforward extension of the algorithm to the 2D case, the image deblurring problem. Section 6 contains numerical results, and section 7 summarizes conclusions and future work.

2. Multilevel algorithm. We first describe the main characteristics of a multilevel method as it applies to a discrete ill-posed problem. The overall idea follows a classical MG approach [5], such as was also used in [6]. Then, we give some details of how we will move between grids.

2.1. Basic framework of a multilevel method. The main idea of a multilevel method is to define a sequence of systems of equations decreasing in size,

$$A^i x^i = b^i, \quad 0 \leq i \leq n,$$

where the superscript i denotes the i th level we have processed. In particular, $i = 0$ corresponds to the finest level, and $i = n$ to the coarsest level. We call these systems (for $i > 0$) *coarse-grid equations*. The restriction operator, R , and the interpolation operator, P , define the transfers from finer to coarser grids and vice versa. The operator at each level is defined by

$$A^{i+1} = R^i A^i P^i.$$

Then, the solution process consists of computing a regularized solution to the fine-scale system (pre-smoothing), computing the residual, transferring the residual from

fine to coarser grid, solving (in the regularized sense) the coarsest-grid correction equation (coarsest-grid correction), interpolation of corrections from the coarse grid, and computing a regularized solution to the updated residual equation on the finer grid (residual correction). The residual-correction step corresponds to post-smoothing in the MG literature. We can summarize the method recursively as follows.

ALGORITHM 1 (Multilevel V-Cycle: MGM).

Input: A^i, b^i

Output: x^i

1. If $i = n$
2. $\text{RegularizedSolve } (A^i x^i = b^i) \quad (\text{Coarsest-grid correction})$
3. Else
4. If $i \neq 0$
5. $\text{RegularizedSolve } (A^i x_{pre}^i = b^i) \quad (\text{Presmoothing})$
6. Else
7. $x_{pre}^0 = 0$
8. End If
9. $r^i = b^i - A^i x_{pre}^i$
10. $b^{i+1} = R^i r^i$
11. $A^{i+1} = R^i A^i P^i$
12. $y^{i+1} = \text{MGM}(A^{i+1}, b^{i+1})$
13. $x^i = x_{pre}^i + P^i y^{i+1}$
14. $r^i = b^i - A^i x^i$
15. $\text{RegularizedSolve } (A^i x_c^i = r^i) \quad (\text{Residual correction})$
16. $x^i = x^i + x_c^i$
17. End If

As in [6], we do not pre-smooth on the finest level. The reason is mainly due to computational cost, since pre-smoothing at the finest level will in fact improve the solution given by this multilevel approach. In order to give a complete definition of a multilevel method, we need to define the restriction and interpolation operators, and also what we mean by pre-smoothing, coarsest-grid correction, and residual correction; that is, we need to define how we solve the systems on lines 2, 5, and 15. We address these issues in the next section.

2.2. Haar decomposition. Recall that, in our problem, A is an $m \times m$ matrix with $m = 2^k$. Consider the Haar matrix transform defined by

$$(2.1) \quad W^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \dots & 1 & 1 \\ 1 & -1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \dots & 1 & -1 \end{pmatrix} = \begin{bmatrix} W_1^T \\ W_2^T \end{bmatrix}.$$

Our problem $Ax = b$ can be written in the wavelet domain as

$$\hat{A}\hat{x} = \hat{b}.$$

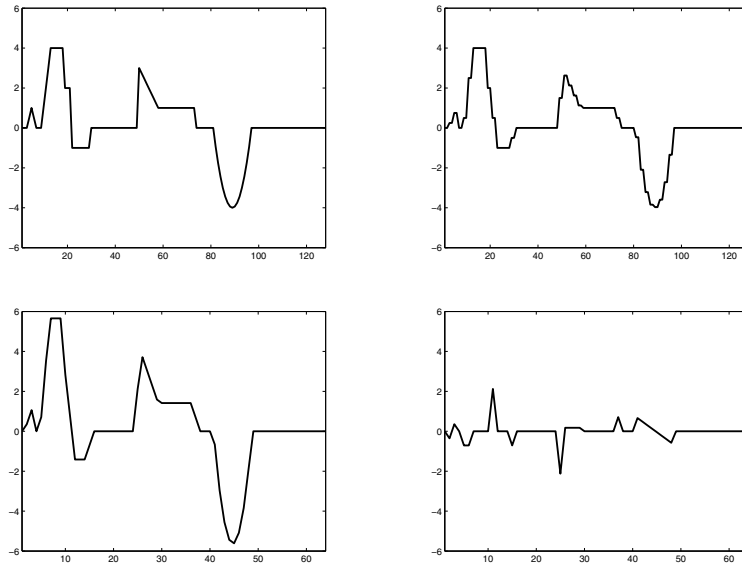


FIG. 2.1. True solution x^{true} (top left), approximate solution $\hat{x} = W[\hat{x}_1; 0]$ (top right), scaling coefficients \hat{x}_1 (bottom left), and wavelet coefficients \hat{x}_2 (bottom right).

In other words, if W^T is the Haar matrix transform,¹ then $\hat{A} = W^T A W$, $\hat{x} = W^T x$, and $\hat{b} = W^T b$.

We partition the transformed problem into blocks of size 2^{k-1} to obtain the following problem:

$$(2.2) \quad \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix},$$

with $\hat{A}_{ij} = W_i^T A W_j$, $\hat{x}_i = W_i^T x$, and $\hat{b}_i = W_i^T b$ for $i, j = 1, 2$.

We now discuss how to use this partitioning to our advantage. It is clear from the definition that W_1^T applied to a vector does weighted averaging, while W_2^T does weighted differencing. Indeed, in the wavelet literature [4] if v is a vector, then $W_1^T v$ is called the vector of *scaling coefficients* and $W_2^T v$ is called the vector of *wavelet coefficients*. In the interest of space, we will not give detailed background on wavelets here, but refer the interested reader to [4] for why Haar wavelets are useful for decomposing signals with edges. We present only the facts most relevant to justifying the use of the Haar basis in our multilevel approach.

Given the preceding description, not surprisingly we observe from Figure 2.1 that the plot of the scaling coefficient vector \hat{x}_1 captures a level of detail similar to the original image, while the plot of the wavelet coefficient vector \hat{x}_2 is spiky. This leads us naturally to say that if we could recover the “downsampled” signal \hat{x}_1 , we would have a rather good approximation to the fine-scale vector (in the sense that setting \hat{x}_2 to 0 and $\hat{x} = W[\hat{x}_1; 0]$ (see Figure 2.1) would already give an okay (but blocky) representation of the signal). So our first goal is to try to recover \hat{x}_1 .

¹We use W^T rather than W for ease in reading the linear algebraic equation.

From the first block equation of (2.2), we have

$$(2.3) \quad \hat{A}_{11}\hat{x}_1 = \hat{b}_1 - \hat{A}_{12}\hat{x}_2.$$

We do not know \hat{x}_2 , of course. In the next section, we discuss why it is appropriate to find a *regularized solution* to

$$(2.4) \quad \hat{A}_{11}\hat{x}_1 = \hat{b}_1$$

for our pre-smoothing and coarsest-grid correction problems, and we discuss what regularization methods are appropriate. Then, we will discuss how to recover \hat{x}_2 , given our estimates of \hat{x}_1 .

3. Analysis and algorithm details. Let us compare the submatrices \hat{A}_{ij} . Recall that A is a discretized version of a blurring operator. Let

$$A = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

be the singular value decomposition (SVD) of A . The matrices U_i and V_i with $i = 1, 2$ have size $2^k \times 2^{k-1}$, and the matrices Σ_i with $i = 1, 2$ have size $2^{k-1} \times 2^{k-1}$. The singular values of A , σ_i , are decreasing from 1 (assuming A has been normalized) to a very small number, with no gap. It is well known that the vectors in U and V increase in frequency from left to right [11], so we can think of the vectors in U_1 and V_1 as the lower frequency vectors, while U_2 and V_2 are the higher frequency. Then

$$(3.1) \quad \hat{A} = \begin{bmatrix} W_1^T U_1 & W_1^T U_2 \\ W_2^T U_1 & W_2^T U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T W_1 & V_1^T W_2 \\ V_2^T W_1 & V_2^T W_2 \end{bmatrix}.$$

If we look at the columns of $W_1^T U_1$, we find low frequency vectors due to the fact that U_1 contains low frequency column vectors and W_1^T smooths them even more. Since we know that the columns of U_1 are smoother than the columns of U_2 , the matrix $W_1^T U_1$ contains vectors with lower frequency than those of $W_1^T U_2$. As we mentioned before, W_2^T acts as a weighted difference. Then, the columns of $W_2^T U_1$ and $W_2^T U_2$ are higher frequency vectors. Thus, $W_1^T U_1$ retains the character of the lowest frequency vectors of U . A similar analysis can be done for the blocks on the rightmost matrix of (3.1), and we therefore conclude that $V_1 W_1^T$ best represents, among all four subblocks, the characteristics of the lowest frequency vectors of V . In particular, we can compute

$$(3.2) \quad \hat{A}_{11} = W_1^T U_1 \Sigma_1 V_1^T W_1 + W_1^T U_2 \Sigma_2 V_2^T W_1,$$

which can be written as $\hat{A}_{11} = W_1^T U_1 \Sigma_1 V_1^T W_1 + E$. We have the following lemma.

LEMMA 3.1. *Given the partitioning above,*

$$|\sigma_s(\hat{A}_{11}) - \sigma_s(W_1^T U_1 \Sigma_1 V_1^T W_1)| \leq \sigma_{2^{k-1}+1}, \quad s = 1, \dots, 2^{k-1},$$

where the notation $\sigma_s(B)$ means the s -largest singular value of the argument B .

Proof. Since W, U, V are matrices with orthonormal columns, it is easy to show that $\|W_i^T U_i\|_2 = 1$ and $\|V_i^T W_i\|_2 = 1$ for $i = 1, 2$. Therefore, $\|E\|_2 = \sigma_{2^{k-1}+1}$. By Corollary 8.6.2 in [7], we know that $|\sigma_s(\hat{A}_{11}) - \sigma_s(W_1^T U_1 \Sigma_1 V_1^T W_1)| \leq \|E\|_2$. Then, the result of the lemma follows. \square

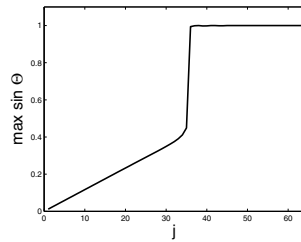


FIG. 3.1. *Maximum sine of the canonical angles between the subspaces spanned by T_1 and T_2 for $j = 1, \dots, 2^{k-1}$.*

Typically, in these problems, the singular value $\sigma_{2^{k-1}+1}$ of A is small, so we can say that the term $W_1^T U_1 \Sigma_1 V_1^T W_1$ is dominant in (3.2). This matrix is almost already in its SVD form, as the matrices $W_1^T U_1$ and $V_1^T W_1$ are close to being orthogonal, and so if Σ_1 is somewhat ill-conditioned, it follows that \hat{A}_{11} inherits these characteristics as well. Specifically, let $\hat{A}_{11} = QDP^T$ be the SVD of \hat{A}_{11} . Let $T_1 = Q_{1:j}$ denote the first j columns of Q , and $T_2 = (W_1^T U_1)_{1:j}$ denote the first j columns of that matrix. Not surprisingly, since the first few (orthonormal) columns of U_1 were smooth and $W_1^T U_1$ is just the averaged version of U_1 , the sine of the canonical angles between the subspaces spanned by T_1 and T_2 has to be small (where small is relative to the size of j). (A similar statement holds for the left singular vectors.) This is illustrated in Figure 3.1 for the blurring operator in our numerical results section. By Corollary I.5.4 in [24], we can define

$$\cos \Theta(\text{Range}(T_1), \text{Range}(T_2)) = \text{diag}(\tilde{\sigma}_1, \dots, \tilde{\sigma}_j),$$

where $\tilde{\sigma}_l$ are the singular values of $T_1^T T_2$. Then,

$$\sin \Theta = \sqrt{1 - \tilde{\sigma}_l^2}, \quad l = 1, \dots, j.$$

In Figure 3.1, we plot the maximum sine of the canonical angles as a function of j .

If we extend this analysis for the operator A^i at each level, then we expect \hat{A}_{11}^i to still be somewhat ill-conditioned on the finest grids (small i), but getting better conditioned the coarser we go (increasing i), and \hat{A}_{11}^i always captures the lowest frequency characteristics of the original matrix A^0 .

3.1. Presmoothing and coarsest-grid correction. It is well known that Krylov subspace algorithms such as LSQR [20] work as regularization methods if stopped early (i.e., before converging to the solution of the system) [11]. This is because these methods, when applied to a discrete ill-posed problem, try to reduce the residual in the first few iterations over the subspace corresponding to the largest singular values, and these are precisely the smooth modes [16]. In other words, if r denotes the residual vector after k steps of LSQR are applied to a discrete ill-posed problem with white noise, a plot of the components of $U^T r$ shows that the first few have been decreased (compared to the initial residual), while the last few have remained relatively untouched (meaning noise has not been mixed into the solution), as shown in Figure 3.2.

The matrix analysis in the previous section has the following implication for our multigrid approach. Suppose we do two or three steps of LSQR on $Ax = b$, form

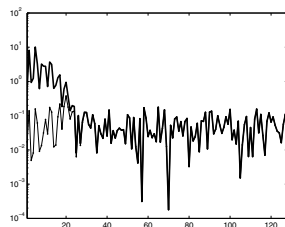


FIG. 3.2. Spectral coefficients of b (thicker line) and spectral coefficients of $r = b - Ax$ (thinner line), where x is given by 3 iterations of LSQR.

$r = b - Ax$, and then form the residual equation in wavelet space:

$$\begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \end{bmatrix}.$$

Consider the system

$$\hat{A}_{11}\hat{x}_1 = \hat{r}_1.$$

Looking at $Q^T \hat{r}_1$, where Q is the matrix of left singular vectors of \hat{A}_{11} , this is equal to

$$Q^T W_1^T r = Q^T W_1^T U U^T r = [Q^T W_1^T U_1, Q^T W_1^T U_2](U^T r).$$

Recall from the previous section that $Q_{1:j}^T(W_1^T U_1)_{1:j} \approx I$. So at least the first j components of this vector are the same in magnitude as they were before—that is, if they were made small on a previous grid, they remain so.

We will consider the term $\hat{A}_{12}\hat{x}_2$ that appears in (2.3) as noise and use LSQR to find an (overregularized) solution to the system

$$(3.3) \quad \hat{A}_{11}\hat{x}_1 = \hat{b}_1,$$

where we use b instead of r for consistency with the previous section. We can consider the term $\hat{A}_{12}\hat{x}_2$ as noise for the following reason. Recall that \hat{x}_2 is the “spiky” wavelet coefficient vector. Applying the operator \hat{A}_{12} smoothes \hat{x}_2 a little (by adding a slightly larger low frequency component to it) but not much because \hat{A}_{12} is not quite a smoother like \hat{A}_{11} is. We also know that $\hat{b}_1 = W_1^T(b^{true} + e) = \hat{b}_1^{true} + \hat{e}_1$. Then, $\hat{A}_{12}\hat{x}_2$ is not white noise, but spectrally it is below the additive white noise, \hat{e}_1 , in our applications, so we can ignore it. Figure 3.3 shows the spectral coordinates of \hat{b}_1^{true} , \hat{e}_1 , and $\hat{A}_{12}\hat{x}_2$ for comparison. By discarding the term $\hat{A}_{12}\hat{x}_2$, a method based on solving (3.3) cannot compute the exact solution of the noise-free problem (1.2).

Again, this analysis is extended to any level. We apply a few iterations of LSQR to obtain an (overregularized) solution, x_{pre}^i , to any intermediate system $A^i x^i = b^i$. At the coarsest level ($i = n$), though, while we may have adequately reduced the residual at the smoothest modes, we may not have obtained as much information over the higher frequency modes of that level’s operator. (Some of the highest frequencies relative to the finest level have, of course, disappeared by this level.) We see that if the original signal has edges, the coarsest-grid problem may still have edges, since

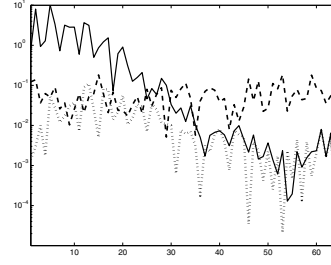


FIG. 3.3. The spectral coordinates of \hat{b}_1^{true} (solid line), \hat{e}_1 (dashed line), and $\hat{A}_{12}\hat{x}_2$ (dotted).

this is the system for the scaling coefficients. To recover that information here, we solve a regularized problem:

$$(3.4) \quad \min_{\hat{x}_1^n} \left\{ \|\hat{A}_{11}^n \hat{x}_1^n - \hat{b}_1^n\|_2^2 + \lambda^q \|L(\hat{x}_1^n + \hat{x}_0^n)\|_q^q \right\},$$

where L is a discrete first-derivative operator, $1 < q < 2$, and \hat{x}_0^n is the initial guess, at the n th level, which is the corresponding projection of the original initial guess x_0^0 and the successive pre-smoothing solutions x_{pre}^i with $i = 1, \dots, n-1$. This allows us to recover some edge information on this grid, but, since we are at a coarse level, this is a much more computationally tractable problem than if we had tried to employ it on the finest grid. On the other hand, if we are at a coarse enough grid, we can replace this simply by a direct solve, as there are no longer any sufficiently small singular values to magnify any remaining noise in the right-hand side (and, in practice, the noise is barely perceptible in the right-hand side at this level).

In summary, we have the following:

- Presmoothing: take a few iterations of LSQR on $A^i x^i = b^i$, $i \neq n$;
- Coarsest-grid correction: if \hat{A}_{11}^n is somewhat ill-conditioned, solve (3.4); otherwise, solve $\hat{A}_{11}^n \hat{x}_1^n = \hat{b}_1^n$ exactly by a direct or iterative method.

The fact that the Toeplitz structure is preserved allows both these steps to be computed efficiently, as we discuss in section 4.1.

3.2. Residual correction. While the above steps can lead to good recovery of the scaling coefficients in the wavelet expansion, we also need to recover the wavelet coefficients \hat{x}_2 in order to adequately recover edges but not produce a solution that looks too artificially blocky. For this part, we define a residual-correction step. Assuming that we have an approximation \hat{x}_1^* of \hat{x}_1 from the coarse-scale (3.3), then we can rewrite system (2.2) as follows:

$$(3.5) \quad \begin{bmatrix} \hat{A}_{12} \\ \hat{A}_{22} \end{bmatrix} \hat{x}_2 = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} - \begin{bmatrix} \hat{A}_{11} \\ \hat{A}_{21} \end{bmatrix} \hat{x}_1^*.$$

We denote the right-hand side of (3.5) by \hat{r}_{new} . To solve this problem, we again use the following Tikhonov formulation to keep a certain smoothness in the solution:

$$(3.6) \quad \min_{\hat{x}_2} \left\{ \left\| \begin{bmatrix} \hat{A}_{12} \\ \hat{A}_{22} \end{bmatrix} \hat{x}_2 - \hat{r}_{new} \right\|_2^2 + \lambda^q \|L(x_{new} + W_2 \hat{x}_2)\|_q^q \right\},$$

where L is a first-derivative operator, $1 < q < 2$ as in (3.4), and $x_{new} = x_0 + x_{pre} + W_1 \hat{x}_1^*$, where x_0 is the initial guess before pre-smoothing on the current level, x_{pre} is

the correction given by the pre-smoothing step, and $W_1\hat{x}_1^*$ is the prolongation of the coarse-scale solution \hat{x}_1^* .

The superscript i is omitted to simplify notation; however, this step is applied to each level. From a computational perspective, it is very important to note that the regularization is applied to the reconstructed signal.

3.3. Our algorithm. Our multilevel algorithm is as follows.

ALGORITHM 2 (Multilevel V-Cycle: MGM).

Input: A^i, b^i, x_0^i

Output: x^i

1. If $i = n$
2. $x^i = \text{direct solve or Newton}(A^i, b^i, x_0^i)$ (see section 4.1)
3. Else
4. If $i \neq 0$
5. $x_{pre}^i = \text{LSQR}(A^i, b^i, \text{iter})$
6. Else
7. $x_{pre}^i = 0$
8. End If
9. $r^i = b^i - A^i x_{pre}^i$
10. $\hat{b}^{i+1} = W_1^T r^i$
11. $\hat{x}_0^{i+1} = W_1^T (x_0^i + x_{pre}^i)$
12. $\hat{A}^{i+1} = W_1^T A^i W_1$ (see section 4)
13. $\hat{x}_1^{i+1} = \text{MGM}(\hat{A}^{i+1}, \hat{b}^{i+1}, \hat{x}_0^{i+1})$
14. $x_{new}^i = x_{pre}^i + W_1 \hat{x}_1^{i+1}$
15. $r_{new}^i = b^i - A^i x_{new}^i$
16. $\hat{x}_2^{i+1} = \text{Newton}(W^T A^i W_2, W^T r_{new}^i, x_0^i + x_{new}^i)$ (see section 4.2)
17. $x^i = x_{new}^i + W_2 \hat{x}_2^{i+1}$
18. End If

Here “iter” in line 5 is the number of iterations of LSQR. We assume $x_0^0 = 0$. A different initial guess for x_0^0 can be added in line 7. Notice that the value of x_0^i is used only for the regularization term.

4. Computational issues. One nice feature of the Haar decomposition is that it preserves the Toeplitz matrix structure. This implies that we can have fast matrix-vector products of the intermediate-level interior linear least squares problems that must be solved iteratively (see next two subsections). In [18], the discrete wavelet transform (DWT) of Toeplitz matrices is discussed. We restrict our study to the Haar wavelet transform (HWT), however, and apply a technique different than that in [18] to prove the following theorem.

THEOREM 4.1. *Let A be an $m \times m$ matrix with Toeplitz structure, and $m = 2^k$. Then, the $2^{k-1} \times 2^{k-1}$ submatrices defined in (2.2), \hat{A}_{11} , \hat{A}_{12} , \hat{A}_{21} , and \hat{A}_{22} , also have a Toeplitz structure.*

Proof. We shall use the representation of a Toeplitz matrix introduced in [15]. Consider the following Toeplitz matrix:

$$A = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \dots & t_{-(m-1)} \\ t_1 & t_0 & t_{-1} & \dots & t_{-(m-2)} \\ t_2 & t_1 & t_0 & \dots & t_{-(m-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{m-1} & t_{m-2} & t_{m-3} & \dots & t_0 \end{bmatrix}.$$

The vector $t = (t_{-(m-1)}, \dots, t_{-1}, t_0, t_1, \dots, t_{m-1})$ is called the *Toeplitz vector* that generates A . If Z_m is the $(2m-1) \times (2m-1)$ “downshift” matrix (i.e., a Toeplitz matrix where its Toeplitz vector t is a vector of all zeros except $t_{-1} = 1$) and P_m is the $m \times (2m-1)$ matrix $[I_m, 0]$, then

$$(4.1) \quad A = [P_m Z_m^{m-1} t, \dots, P_m Z_m^1 t, P_m Z_m^0 t].$$

We know that $\text{vec}(W^T A W) = (W^T \otimes W^T) \text{vec}(A)$ (see Lemma 8.5.3 in [3]), where $E \otimes F$ denotes the Kronecker product of E and F [3, section 8.5], and $\text{vec}(E)$ unstacks matrix E by columns to create a column vector. Then,

$$\begin{aligned} \text{vec}(\hat{A}) &= (W^T \otimes W^T) \text{vec}(A) = (W^T \otimes I) \begin{bmatrix} W^T P_m Z_m^{m-1} t \\ \vdots \\ W^T P_m Z_m^0 t \end{bmatrix} = (W^T \otimes I) \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} c_1 + c_2 \\ c_3 + c_4 \\ \vdots \\ c_{m-1} + c_m \\ c_1 - c_2 \\ c_3 - c_4 \\ \vdots \\ c_{m-1} - c_m \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} W^T P_m Z_m^{m-2} (Z_m^1 + Z_m^0) t \\ W^T P_m Z_m^{m-4} (Z_m^1 + Z_m^0) t \\ \vdots \\ W^T P_m Z_m^0 (Z_m^1 + Z_m^0) t \\ W^T P_m Z_m^{m-2} (Z_m^1 - Z_m^0) t \\ W^T P_m Z_m^{m-4} (Z_m^1 - Z_m^0) t \\ \vdots \\ W^T P_m Z_m^0 (Z_m^1 - Z_m^0) t \end{bmatrix}. \end{aligned}$$

Define $\bar{t} = (Z_m + I)t$ and $\tilde{t} = (Z_m - I)t$. It follows that

$$\hat{A} = \frac{1}{\sqrt{2}} W^T P_m [Z_m^{m-2} \bar{t}, Z_m^{m-4} \bar{t}, \dots, \bar{t}, Z_m^{m-2} \tilde{t}, Z_m^{m-4} \tilde{t}, \dots, \tilde{t}].$$

Then, we have

$$(4.2) \quad \hat{A} = \frac{1}{2} \left[\begin{array}{c|c} P_{m/2} Z_{m/2}^{\frac{m}{2}-1} \bar{t}_{11}, \dots, P_{m/2} Z_{m/2}^0 \bar{t}_{11} & P_{m/2} Z_{m/2}^{\frac{m}{2}-1} \bar{t}_{12}, \dots, P_{m/2} Z_{m/2}^0 \bar{t}_{12} \\ \hline P_{m/2} Z_{m/2}^{\frac{m}{2}-1} \tilde{t}_{21}, \dots, P_{m/2} Z_{m/2}^0 \tilde{t}_{21} & P_{m/2} Z_{m/2}^{\frac{m}{2}-1} \tilde{t}_{22}, \dots, P_{m/2} Z_{m/2}^0 \tilde{t}_{22} \end{array} \right],$$

where the vectors \bar{t}_{11} , \bar{t}_{12} , \bar{t}_{21} , and \bar{t}_{22} are defined (using MATLAB notation) as follows:

$$\begin{aligned} t_{11} &= (I + Z_m) \bar{t}, & t_{12} &= (I + Z_m) \tilde{t}, & t_{21} &= (I - Z_m) \bar{t}, & t_{22} &= (I - Z_m) \tilde{t}, \\ \bar{t}_{11} &= t_{11}(1:2:2m-3), & \bar{t}_{12} &= t_{12}(1:2:2m-3), & \bar{t}_{21} &= t_{21}(1:2:2m-3), & \bar{t}_{22} &= t_{22}(1:2:2m-3). \end{aligned}$$

From (4.2) and from the characterization of Toeplitz matrices (4.1), the result of the theorem follows. Moreover, we have shown that the vectors $\frac{1}{2} \bar{t}_{11}$, $\frac{1}{2} \bar{t}_{12}$, $\frac{1}{2} \bar{t}_{21}$, and $\frac{1}{2} \bar{t}_{22}$ are their respective Toeplitz vectors. \square

There are two important computational issues deduced from Theorem 4.1:

- we can compute the entries of \hat{A}_{ij} with $i, j = 1, 2$ without explicitly computing the matrices;
- since \hat{A}_{ij} are Toeplitz, they admit fast matrix-vector products. This is a computational advantage for LSQR on the downward part of the V-cycle and also for solving the system or regularized problem on the coarsest level (as noted below).

Furthermore, we can show that the bandwidth of the Toeplitz matrices at each grid level is reduced. Then, the coarsest-grid operator can be reduced as far as to a tridiagonal matrix, depending on the initial bandwidth and how many levels are used, again reducing the overall computational burden further down the V-cycle.

THEOREM 4.2. *If A is an $m \times m$, $m = 2^k$, banded Toeplitz matrix with upper bandwidth k_u and lower bandwidth k_l , then \hat{A}_{ij} are $m/2 \times m/2$ banded Toeplitz matrices with bandwidths $\lceil k_u/2 \rceil$ and $\lceil k_l/2 \rceil$.*

Proof. As in the preceding proof, we deal directly with the Toeplitz vectors. From the proof above, we know that Toeplitz vectors that generate \hat{A}_{ij} can be written as $\bar{t}_{ij} = t_{ij}(1 : 2 : 2m - 3)$, where t_{ij} is a $2m - 1$ length vector $t_{ij} = \alpha It + \beta Z_m t + \gamma Z_m^2 t$ with α and γ either plus or minus 1 and β either 0 or 2.

Thus each t_{ij} is easily shown to have $m - k_u - 3$ leading zeros, followed by $k_l + k_u + 3$ nonzeros, followed by $m - k_l - 1$ trailing zeros. Since \bar{t}_{ij} is obtained by subsampling t_{ij} in every other entry starting with the first, there must be $\lceil \frac{m - k_u - 3}{2} \rceil$ leading zeros in \bar{t}_{ij} . Clearly, the upper bandwidth of the $m/2 \times m/2$ matrices \hat{A}_{ij} is obtained by subtracting the number of leading zeros plus 1 (to account for the main diagonal) from $m/2$. Given the number of nonzeros in the “middle” of the Toeplitz vector \bar{t}_{ij} and the upper bandwidth, we can compute the lower bandwidth.

Case 1: $k_u = 2s + 1$ for positive integer s . The upper bandwidth is easily shown to be $s + 1$ (equivalently, $\lceil \frac{k_u}{2} \rceil$). Since k_u is odd, given the subsampling pattern, there must be $\lceil \frac{k_l + k_u + 3}{2} \rceil = s + 2 + \lceil \frac{k_l}{2} \rceil$ nonzeros after the leading zeros in \bar{t}_{ij} . Hence, the lower bandwidth of \hat{A}_{ij} is $s + 2 + \lceil \frac{k_l}{2} \rceil - 1 - (s + 1) = \lceil \frac{k_l}{2} \rceil$.

Case 2: $k_u = 2s$, which implies that the upper bandwidth of each \hat{A}_{ij} is $\frac{m}{2} - (\frac{m}{2} - s - 1) - 1 = s = \frac{k_u}{2}$. Since k_u is even, there are $\lfloor \frac{k_l + k_u + 3}{2} \rfloor = s + \lfloor \frac{k_l + 3}{2} \rfloor$ nonzeros after the leading zeros in \bar{t}_{ij} . Thus the lower bandwidth is $s + \lfloor \frac{k_l + 3}{2} \rfloor - 1 - s$, which is easily shown to be $\lceil \frac{k_l}{2} \rceil$ regardless of whether k_l is even or odd. \square

4.1. Coarse-scale solve and parameter issue. Recall that we may need to solve (3.4) on the coarsest grid if not enough levels are used. If $q = 2$, we could use a hybrid approach [14] to solve the Tikhonov problem and simultaneously select the regularization parameter in an efficient manner. However, we typically take $1 < q < 2$, with $q \ll 2$ to best preserve the edges at the coarsest grid. To solve the minimization problem (3.4), we use Newton’s method with line search. The recursion of this method is defined by $x_{k+1} = x_k + \alpha_k d_k$, where α_k is the length of the step² and d_k is its direction. Recall that if we minimize a function $s(x)$, then the direction d is the solution to the linear system

$$(4.3) \quad H(x)d = -\nabla s(x),$$

where $\nabla s(x)$ is the gradient of s with respect to x , and $H(x)$ is the *Hessian* matrix, containing second derivatives: $H_{ij} = \partial^2 s(x) / \partial x_i \partial x_j$. In our particular case, we have that

$$s(\hat{x}_1) = \|\hat{A}_{11}\hat{x}_1 - \hat{b}_1\|_2^2 + \lambda^q \|L(\hat{x}_1 + \hat{x}_0)\|_q^q.$$

Then, the corresponding gradient and Hessian are as follows:

$$\nabla s(\hat{x}_1) = 2\hat{A}_{11}^T \bar{r}_1 + \lambda^q q L^T \text{sign}(\bar{r}_2) \cdot |\bar{r}_2|^{q-1}$$

²The value of α_k is chosen using a standard line search.

and

$$H(\hat{x}_1) = 2\hat{A}_{11}^T \hat{A}_{11} + \lambda^q q(q-1)L^T \text{diag}(|\bar{r}_2|^{q-2})L,$$

where $\bar{r}_1 = \hat{A}_{11}\hat{x}_1 - \hat{b}_1$, $\bar{r}_2 = L(\hat{x}_1 + \hat{x}_0)$ and, for a vector v , $\text{sign}(v)$ and $|v|$ are the vectors containing the sign and absolute value of each entry of v , and $\text{diag}(v)$ is the diagonal matrix with entries given by v ; also for two vectors v and w , $v.*w$ is their coordinatewise product. Then, the system of (4.3) can be written as

$$(4.4) \quad (\hat{A}_{11}^T \hat{A}_{11} + L^T D L)d = - \left(\hat{A}_{11}^T \bar{r}_1 + \frac{1}{q-1} L^T D \bar{r}_2 \right),$$

where $D = \frac{\lambda^q}{2} q(q-1) \text{diag}(|\bar{r}_2|^{q-2})$. To solve (4.4), we can apply an iterative method such as CG or MINRES [7], or we can also solve its equivalent linear least squares problem,

$$(4.5) \quad \min_d \left\| \begin{bmatrix} \hat{A}_{11} \\ D^{1/2} L \end{bmatrix} d - \begin{bmatrix} -\bar{r}_1 \\ -\frac{1}{q-1} D^{1/2} \bar{r}_2 \end{bmatrix} \right\|_2^2,$$

by applying an iterative method such as LSQR or CGLS that needs only the matrix vector product $\hat{A}_{11}v$ and $D^{1/2}Lv$.

Choosing a good regularization parameter specifically for (3.4) is beyond the scope of this paper. In our numerical section, we present results for which λ was chosen to be optimal over a small, discrete set of choices. Choosing the parameter at this level is the subject of current research.

4.2. Residual correction solve. To solve the minimization problem (3.6), we also use Newton's method with line search. In this case, we have that

$$s(\hat{x}_2) = \|\tilde{r}_1\|_2^2 + \|\tilde{r}_2\|_q^q,$$

where

$$\tilde{r}_1 = \begin{bmatrix} \hat{A}_{12} \\ \hat{A}_{22} \end{bmatrix} \hat{x}_2 - \hat{r}_{new} \quad \text{and} \quad \tilde{r}_2 = \lambda(Lx_{new} + LW_2 \hat{x}_2).$$

Then, the corresponding gradient and Hessian are as follows:

$$\nabla s(\hat{x}_2) = 2 \begin{bmatrix} \hat{A}_{12} \\ \hat{A}_{22} \end{bmatrix}^T \tilde{r}_1 + q(\lambda LW_2)^T \text{sign}(\tilde{r}_2). * |\tilde{r}_2|^{q-1}$$

and

$$H(\hat{x}_2) = 2 \begin{bmatrix} \hat{A}_{12} \\ \hat{A}_{22} \end{bmatrix}^T \begin{bmatrix} \hat{A}_{12} \\ \hat{A}_{22} \end{bmatrix} + q(q-1)(\lambda LW_2)^T \text{diag}(|\tilde{r}_2|^{q-2})(\lambda LW_2).$$

Then, the system of (4.3) can be written as

$$(4.6) \quad A_L^T D A_L d = A_L^T D w,$$

where

$$A_L = \begin{bmatrix} \hat{A}_{12} \\ \hat{A}_{22} \\ \lambda LW_2 \end{bmatrix}, \quad D = \begin{bmatrix} 2I & 0 \\ 0 & D_q \end{bmatrix}, \quad \text{and} \quad w = \begin{bmatrix} -\tilde{r}_1 \\ -\frac{1}{q-1} \tilde{r}_2 \end{bmatrix},$$

with $D_q = q(q-1)\text{diag}(|\tilde{r}_2|^{q-2})$. To solve (4.6), we can apply an iterative method such as CG or MINRES [7]. We do not need to construct the matrix $A_L^T D A_L$ explicitly. We need only the matrix-vector product. Since D is diagonal and A_L has Toeplitz and sparse blocks, this is efficiently solved using iterative solvers that capitalize on this structure for matrix-vector products. We can also solve its equivalent linear least squares problem,

$$(4.7) \quad \min_d \|D^{1/2} A_L d - D^{1/2} w\|_2^2,$$

by applying an iterative method such as LSQR or CGLS. Also in this case, we do not need to construct the matrix $D^{1/2} A_L$ explicitly.

Recall that \hat{x}_2 lives on a coarser grid than where we are enforcing the regularization (at the next finest grid level). As we have seen in Figure 2.1, the wavelet coefficients are spiky. It is still worthwhile to use q as close to 1 as possible to recover wavelet coefficients accurately enough to represent the edges appropriately on the next finest grid level. As mentioned in the previous subsection, the selection of an optimal regularization parameter is beyond the scope of the present work.

5. The 2D case: The image restoration problem. In this section, we discuss a straightforward extension of the above approach to the image (2D signal) restoration problem by considering a simplified type of 2D blurring kernel. As we note below, there are subtle issues associated with the 2D case that warrant further investigation. However, the results here show the potential for our concept on 2D problems.

A first-kind Fredholm integral equation for a 2D problem is as follows:

$$(5.1) \quad \int_{\alpha_1}^{\alpha_2} \int_{\gamma_1}^{\gamma_2} K(s, s', t, t') f(t, t') dt dt' = g(s, s').$$

We consider the special case where the kernel $K(s, t)$ is spatially invariant and separable, i.e., $K(s, s', t, t') = l(s - t)w(s' - t')$. Moreover, we shall assume for simplicity that $l = w$. If A is the discretization of the integral operator l , then the corresponding discrete operator \mathbf{A} of the kernel K is defined by

$$\mathbf{A} = A \otimes A \in \mathbb{R}^{m^2 \times m^2},$$

with $m = 2^k$, where the matrix \mathbf{A} acts as a *blurring* operator on the image x^{true} and b denotes the blurred noisy image. We can state our image reconstruction problem as finding x such that

$$\mathbf{A}x = b,$$

where x contains the elements of the 2D image by stacking its columns.

The 2D HWT $\tilde{\mathbf{W}}$ can be defined as a Kronecker product of the 1D HWT W , that is, $\tilde{\mathbf{W}} = W \otimes W$. To get a convenient matrix representation we use $\mathbf{W} = P\tilde{\mathbf{W}}$, where P is specially defined so that the system in the wavelet domain resembles the 1D version of the problem in the wavelet domain. That is, the system in the wavelet domain has the form

$$\begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ \hat{\mathbf{A}}_{21} & \hat{\mathbf{A}}_{22} \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix}, \quad \tilde{x}_1 = \hat{x}_{11} = (W_1 \otimes W_1)x,$$

$$\tilde{x}_2 = \begin{bmatrix} \hat{x}_{21} \\ \hat{x}_{12} \\ \hat{x}_{22} \end{bmatrix} = \begin{bmatrix} (W_2 \otimes W_1)x \\ (W_1 \otimes W_2)x \\ (W_2 \otimes W_2)x \end{bmatrix},$$

with similar definitions for \tilde{b}_1 and \tilde{b}_2 . Note that, like the 1D case, \hat{x}_{11} contains pure scaling coefficients, but unlike the 1D case, the system is not split in half.

From the first equation we have

$$\hat{\mathbf{A}}_{11}\tilde{x}_1 + \hat{\mathbf{A}}_{12}\tilde{x}_2 = \tilde{b}_1.$$

As before, we discard $\hat{\mathbf{A}}_{12}\tilde{x}_2$ and solve $\hat{\mathbf{A}}_{11}\tilde{x}_1 = \tilde{b}_1$ to get \tilde{x}_1 . Note, however, that $\hat{\mathbf{A}}_{11} = \hat{A}_{11} \otimes \hat{A}_{11}$. We have shown for the 1D case that \hat{A}_{11} is still ill-conditioned and that it captures the lowest frequency characteristics of A . Thus, we expect the matrix $\hat{\mathbf{A}}_{11}$ to be somewhat ill-conditioned, since its condition number $\kappa(\hat{A}_{11} \otimes \hat{A}_{11})$ (that is, $\|\hat{A}_{11} \otimes \hat{A}_{11}\| \|(\hat{A}_{11} \otimes \hat{A}_{11})^{-1}\|$) is $\kappa(\hat{A}_{11})\kappa(\hat{A}_{11})$, and to capture the lowest frequency characteristics of \mathbf{A} .

The extension to the 2D case seems straightforward by following the same steps used to develop the algorithm for the 1D case. However, if we do a similar spectral analysis to justify why it is allowable to discard $\hat{\mathbf{A}}_{12}\tilde{x}_2$, we find that important information is lost by this model simplification. Furthermore, efficiency is lost in that the system size is not halved each time. We can still compute a regularized solution, but it is possible to obtain a better reconstruction. So a more careful algorithm needs to be developed. Current research aims to address this issue by means of a wavelet-packet-based multilevel approach.

In terms of computational issues, the inheritance of structure from the 1D case is also seen in the 2D case. In the case of the image deblurring problem, the $m^2 \times m^2$ matrix \mathbf{A} is block-Toeplitz–Toeplitz-block (BTTB), that is, an $m \times m$ block-Toeplitz matrix with each block being an $m \times m$ Toeplitz matrix. So is $\hat{\mathbf{A}}_{11}$. Thus, in the 2D case the BTTB structure is kept and can be used for fast matrix multiplication.

6. Numerical results. We conclude with illustrations of the performance of our algorithm in signal and image restoration problems. All computations were performed in MATLAB. In the following examples, we aim to recover the original signal/image from a blurred and noisy signal/image.

6.1. Example 1: 1D signal restoration. The blurring operator is defined by an $m \times m$ symmetric banded Toeplitz matrix A , where its first row is defined by the vector

$$(6.1) \quad z = \frac{1}{2\pi\sigma^2} [\exp(-([0 : \text{band} - 1].^2)/(2\sigma^2)); \text{zeros}(1, m - \text{band})].$$

The parameter band specifies the half-bandwidth of the matrix A , and the parameter σ corresponds to the variance of the Gaussian point spread function,

$$h(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

We use $m = 128$, and the parameters $\sigma = 3$ and $\text{band} = 7$. The condition number of A is $\kappa(A) = 4.8 \times 10^5$. The exact solution, represented by x^{true} , is the vector of length 128 shown in Figure 2.1. The noise-free blurred signal, represented by b^{true} , is computed as $b^{true} = Ax^{true}$. The elements of the noise vector e are normally distributed with zero mean, and the standard deviation is chosen such that $\frac{\|e\|_2}{\|b^{true}\|_2} = 0.05$. In this case, we say that the level of noise is 5%. The noisy right-hand side of our system is defined by $b = b^{true} + e$ (see Figure 6.1).

For the solutions shown in Figure 6.1, we define the initial guess $x_0^0 = 0$, and we apply nine iterations of LSQR on the pre-smoothing step. The operator at the

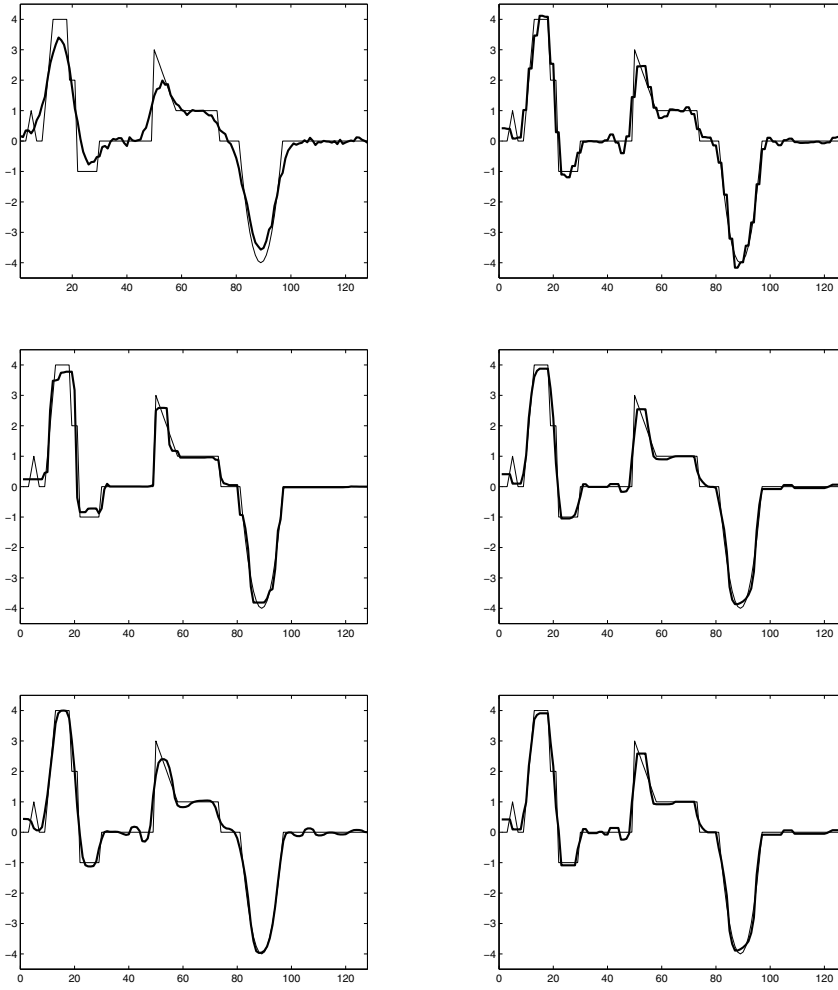


FIG. 6.1. Blurred noisy right-hand side, b (top left), 3-MGM solution without residual-correction step (top right), 3-MGM solution without pre-smoothing step (no LSQR iteration) (middle left), 3-MGM solution with $q = 1.1$ (middle right), 3-MGM solution with $q = 1.5$ (bottom left), and 3-MGM solution with $q = 1.01$. In all figures, the exact solution x^{true} appears as a thin line.

coarsest level has size 32×32 (which means we apply a 3-level MGM), and a Newton method is applied to solve the coarsest-grid correction. For the regularization term in (3.4) and (3.6), we use $q = 1.1$ and the matrix

$$L = \begin{pmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 & 1 \end{pmatrix},$$

which is a scaled discrete approximation to the first derivative operator on a regular grid, with no assumptions on boundary conditions. We define ten logarithmically spaced values of λ from 10^{-3} to 1. We loop for all values of λ at each level and choose the combination of λ values that makes $\frac{\|x^{true} - x^{MGM}(\lambda^{ijk})\|_1}{\|x^{true}\|_1}$ minimum. We test the

TABLE 1
Comparison of different levels and parts of the multilevel method.

Level	LSQR iter	q	Newton iter	λ_{opt}	$\frac{\ x^{true} - x^{MGM}\ _1}{\ x^{true}\ _1}$
1	-	1.1	34	λ^7	0.1821
2	9	1.1	31,29	$\lambda^{4,10}$	0.1963
3	9	1.1	19,27,28	$\lambda^{4,5,10}$	0.1736
3	9	1.1	23,0,0	$\lambda^{4,-,-}$	0.2397
3	0	1.1	23,29,34	$\lambda^{8,6,5}$	0.1902
3	9	1.01	47,90,45	$\lambda^{3,4,10}$	0.1796
3	9	1.5	5,9,7	$\lambda^{5,7,9}$	0.1819

TABLE 2
Comparison of different methods applied to different levels of noise.

Noise level	LSQR(iter)	TV (λ_i)	MGM (iter, λ^{ijk})
1%	0.1742 (24)	0.1637 (λ_2)	0.1340 (24, $\lambda^{5,5,10}$)
5%	0.2165 (8)	0.2065 (λ_4)	0.1736 (9, $\lambda^{4,6,10}$)
10%	0.2191 (6)	0.2187 (λ_7)	0.2049 (2, $\lambda^{7,7,8}$)

method with $q = 1.1$ in both minimization problems (3.6) and (3.5). We observe that using both the pre-smoothing and residual-correction steps is essential to recover the signal. Without pre-smoothing we miss some edge information, and we obtain a blocky solution if we do not use the residual-correction step in our algorithm.

We also test the method for different number of levels; that is, we compute the solution of 1-level, 2-level, and 3-level MGMs. In Table 1, we report the number of iterations for Newton's method at each level, the optimal values of λ for each level, and the relative errors computed by $\frac{\|x^{true} - x^{MGM}\|_1}{\|x^{true}\|_1}$. The leftmost Newton iteration number is the number of iterations on the coarsest grid, and they read from left to right with respect to coarseness of the grid. Recall that for multiple levels, the last Newton iteration is on a residual correction equation where the solution is on the next-to-finest grid and only the regularization is applied at the finest level. The notation λ^{ijk} means that for the coarsest level we need $\lambda(i)$, for level 1 we need $\lambda(j)$, and for the finest level we need $\lambda(k)$.

We also compare solutions for different values of q since the algorithm is sensitive to changes in it. When q is very close to 1, then the number of iterations needed for Newton's method to converge increases. On the other hand, when $q = 1.5$, only a few iterations are needed. However, we can see in Figure 6.1 that, in the case when $q = 1.5$, the solution is smooth, losing some edge information. One could opt to increase q with decreasing coarseness in an attempt to achieve more balance, but we did not investigate that possibility here.

For different levels of noise (1%, 5%, and 10%), we compute an LSQR solution, $x^{LSQR(iter)}$, and a TV solution, $x^{TV}(\lambda)$, at the finest level for comparison. To compute x^{TV} we applied a primal-dual Newton method (see section 8.2.5 in [25]). We define ten logarithmically spaced values of λ from 10^{-4} to 1. The optimal parameters k and λ are chosen such that $\frac{\|x^{true} - x^{LSQR(iter)}\|_1}{\|x^{true}\|_1}$ and $\frac{\|x^{true} - x^{TV}(\lambda)\|_1}{\|x^{true}\|_1}$ are minimums. For a quantitative comparison, see Table 2. For qualitative comparison in the case when the noise level is 10%, see Figure 6.2.

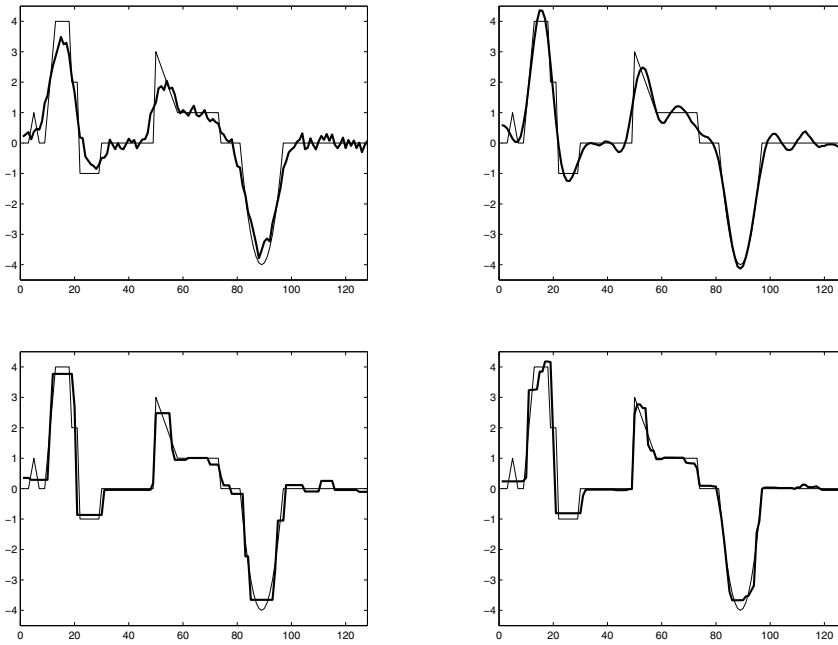


FIG. 6.2. Blurred noisy right-hand side, b (top left), LSQR solution (top right), TV solution (bottom left), MGM solution (bottom right). In all figures, the exact solution x^{true} appears as a thin line.

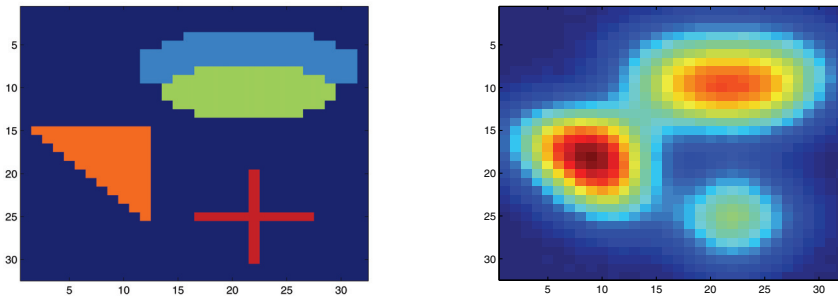
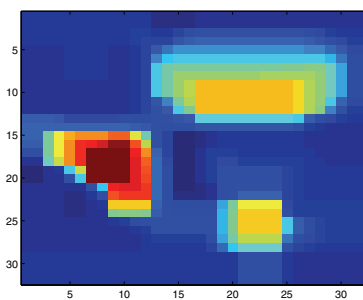
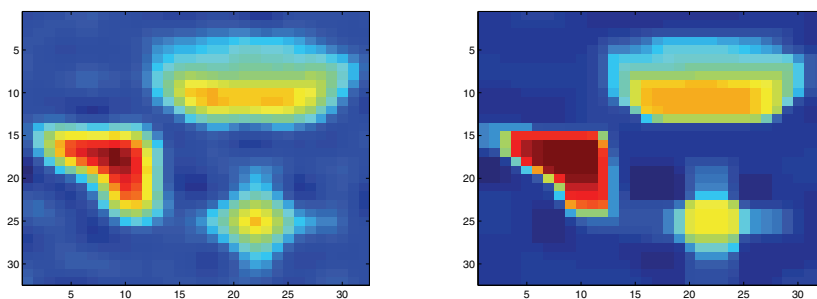


FIG. 6.3. True solution x^{true} (left) and the blurred noisy right-hand side, b (right).

6.2. Example 2: 2D image restoration. Our next example is a 2D image deblurring problem which consists of recovering the original $N \times N$ image from an image blurred by the optical system used to record it. The blur operator is defined by the $N^2 \times N^2$ matrix $\mathbf{A} = (2\pi\sigma^2)A \otimes A$, where A is an $N \times N$ symmetric banded Toeplitz matrix whose first row is also defined by (6.1). We use $N = 32$, and parameters $\sigma = 3$ and band = 9. The matrix \mathbf{A} has a condition number $\kappa(\mathbf{A}) = 3.75 \times 10^{11}$. The original image, obtained from the Regularization Tools package by Hansen [10], is shown in Figure 6.3. It is a 32×32 image that is columnwise stacked in the vector x^{true} . The vector b^{true} , containing the noise-free blurred image, is computed as $b^{true} = \mathbf{A}x^{true}$. The entries of the error vector e are normally distributed with

FIG. 6.4. *The MGM solution.*FIG. 6.5. *Solutions computed at the finest level using LSQR (left) and Newton (1-MGM) (right).*TABLE 3
Comparison of different methods: relative errors.

LSQR (41)	Newton (λ_2)	3-MGM (50, $\lambda^{5,6,8}$)
0.6650	0.4925	0.5459

zero mean and normalized such that $\frac{\|e\|_2}{\|b^{true}\|_2} = 0.01$. The contaminated right-hand side of our system is defined by $b = b^{true} + e$ (see Figure 6.3).

For the 3-MGM solution shown in Figure 6.4, we define the initial guess $x_0^0 = 0$, and we apply 50 iterations of LSQR on the pre-smoothing step. The image at the coarsest level has size 8×8 , and a Newton method is applied to solve the coarsest-grid correction. For the regularization term in (3.4) and (3.6), we use $q = 1.1$ and the matrix

$$L = \begin{bmatrix} I & \otimes & L_1 \\ L_1 & \otimes & I \end{bmatrix},$$

where L_1 is the regularization operator used in Example 1.

We define ten logarithmically spaced values of λ from 10^{-4} to 1. We loop for all values of λ at each level and choose the combination of λ values that makes $\frac{\|x^{true} - x^{MGM}(\lambda^{ijk})\|_1}{\|x^{true}\|_1}$ minimum.

We also compute both $x^{LSQR(iter)}$ and 1-MGM (λ_i) solutions for comparison. The optimal parameters k, λ are chosen such that the 1-norm relative errors are minimums. Table 3 shows optimal parameters and relative errors. Figures 6.4 and 6.5 show the reconstructed images.

7. Conclusion. In this paper, a new edge-preserving multilevel algorithm for signal restoration problems has been presented. The transformation of the problem to the wavelet domain using Haar wavelets leads to a multilevel algorithm that preserves matrix structures. In particular, we have shown that Toeplitz structure is preserved between grids, which leads to intermediate linear system solves inside damped Newton iterations that are structured. This means that iterative methods can be efficiently applied to the linear systems. Numerical experiments illustrate that this is a promising technique for 1D restoration that can be extended in a straightforward way to the 2D case. A detailed study of more appropriate modifications of the method to handle 2D restoration is currently under investigation. Future work includes efficient parameter selection techniques, number of smoothing steps at intermediate grids, and number of V-cycles.

Acknowledgments. We thank Scott MacLachlan and the two anonymous reviewers for their careful reading of the manuscript and for valuable comments and suggestions that helped to improve the paper.

REFERENCES

- [1] A. ARICÒ AND M. DONATELLI, *A V-cycle multigrid for multilevel matrix algebras: Proof of optimality*, Numer. Math., 105 (2007), pp. 511–547.
- [2] A. ARICÒ, M. DONATELLI, AND S. SERRA-CAPIZZANO, *V-cycle optimal convergence for certain (multilevel) structured linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 186–214.
- [3] Å. BJÖRCK, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
- [4] A. BOGGESE AND F. J. NARCOWICH, *First Course in Wavelets with Fourier Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 2001.
- [5] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A Multigrid Tutorial*, 2nd ed., SIAM, Philadelphia, 2000.
- [6] M. DONATELLI AND S. SERRA-CAPIZZANO, *On the regularizing power of multigrid-type algorithms*, SIAM J. Sci. Comput., 27 (2006), pp. 2053–2076.
- [7] G. H. GOLUB AND C. F. V. LOAN, *Matrix Computations*, 3rd. ed., The Johns Hopkins University Press, Baltimore, MD, 1996.
- [8] C. W. GROETSCH, *Inverse Problems in Mathematical Sciences*, Vieweg, Wiesbaden, Germany, 1993.
- [9] M. HANKE AND C. R. VOGEL, *Two-level preconditioners for regularized inverse problems I: Theory*, Numer. Math., 83 (1999), pp. 385–402.
- [10] P. C. HANSEN, *Regularization tools: Matlab package for the analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.
- [11] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problem: Numerical Aspects of Linear Inversion*, SIAM Monogr. Math. Model. Comput. 4, SIAM, Philadelphia, 1998.
- [12] P. C. HANSEN, M. JACOBSEN, J. RASMUSSEN, AND H. SØRENSEN, *The PP-TSVD algorithm for image restoration problems*, in Methods and Applications of Inversion, P. C. Hansen, B. H. Jacobsen, and K. Mosegaard, eds., Lecture Notes in Earth Sci. 92, Springer, Berlin, 2000, pp. 171–186.
- [13] B. KALTENBACHER, *On the regularizing properties of a full multigrid method for ill-posed problems*, Inverse Problems, 17 (2001), pp. 767–788.
- [14] M. E. KILMER, P. C. HANSEN, AND M. I. ESPAÑOL, *A projection-based approach to general-form Tikhonov regularization*, SIAM J. Sci. Comput., 29 (2007), pp. 315–330.
- [15] M. E. KILMER AND J. G. NAGY, *Kronecker product approximations for dense block Toeplitz-plus-Hankel matrices*, Numer. Linear Algebra Appl., 14 (2007), pp. 581–602.
- [16] M. E. KILMER AND G. W. STEWART, *Iterative regularization and MINRES*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 613–628.
- [17] J. T. KING, *Multilevel algorithms for ill-posed problems*, Numer. Math., 61 (1992), pp. 311–334.
- [18] F. R. LIN, W. K. CHING, AND M. K. NG, *Discrete wavelet transforms for Toeplitz matrices*, Linear Algebra Appl., 370 (2003), pp. 269–285.
- [19] S. MORIGI, L. REICHEL, F. SGALLARI, AND A. SHYSHKOV, *Cascadic multiresolution methods for image deblurring*, SIAM J. Imaging Sci., 1 (2008), pp. 51–74.
- [20] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.

- [21] L. REICHEL AND A. SHYSHKOV, *Cascadic multilevel methods for ill-posed problems*, J. Comput. Appl. Math., to appear.
- [22] A. RIEDER, *A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization*, Numer. Math., 75 (1997), pp. 501–522.
- [23] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [24] G. W. STEWART AND J. G. SUN, *Matrix Perturbation Theory*, Academic Press, New York, 1990.
- [25] C. R. VOGEL, *Computational Methods for Inverse Problems*, Frontiers Appl. Math. 23, SIAM, Philadelphia, 2002.